

Measuring Haptic Experience: Elaborating the HX model with scale development

Suji Sathiyamurthy¹, Melody Lui¹, Erin Kim², and Oliver Schneider¹

Abstract—Designers increasingly employ haptic feedback with the aim to improve user experience (UX). Designers and researchers currently use qualitative methods or demos for feedback, but neither approach scales to large studies or remote work. We build upon the recent Haptic Experience (HX) model and report on progress towards measuring the five constructs for designing haptic experiences: Harmony, Expressivity, Autotelics, Immersion, and Realism. We describe initial findings from scale development, specifically, from item generation (N=23) and exploratory factor analysis (N=261). Our results provide initial evidence that vibrotactile experiences are effectively modeled by five factors, enriched description of each factor, and guidelines for quantitatively measuring HX.

I. INTRODUCTION

Haptic technology is becoming an essential tool for designers seeking to create great user experience (UX). There is mounting evidence that haptic feedback of different types can contribute to existing UX measures. Mid-air haptic feedback can make videos measurably more pleasant, unpredictable, and creative [1], vibrotactile feedback can improve player experience [2], and motion seats can evoke better experiences, as measured by EEG and other physiological signals [3]. In virtual reality (VR) environments, well-designed haptic feedback can lead to increased presence [4], [5]. However, these metrics give little insight into how and why haptics contributes to peoples' experience, and little direction for hapticians to improve their designs.

Currently, designers, researchers, and hapticians use qualitative methods to gain deeper insight into their designs. From investigation into haptic experience design [6], we know practitioners favour small, in-person acceptance tests to evaluate their designs, iterating until it just “feels right.” When communicating the efficacy of a design, in-person demos are the best way to persuade stakeholders. Neither approach scales to large quantitative studies or translates to remote work, and while some hapticians employ general scales like the AttrakDiff questionnaire [7], [8], there is still a desire for more formal measurement tools.

In this paper, we report on initial findings from conducting scale development to measure haptic experience (HX) [9]. Scale development – designing and validating a measurement

instrument – is considered critical to building theoretical knowledge in human and social sciences [10]. We employed scale development to better understand the HX model, a proposed standard for defining haptic experience in terms of its pragmatic and hedonic factors [9]. Based on this model, we believe that the 5 experiential factors of “Harmony”, “Autotelics”, “Expressivity”, “Immersion” and “Realism” could be guiding constructs important to measuring HX. We conducted the first steps of scale development to assess how the HX model fits different devices and applications, understand user’s expectations and emotions, and analyze whether the HX model is supported by empirical evidence. Using the results of a series of 5 user studies and exploratory factor analysis (EFA), we developed an initial draft of a measurement instrument and here report on the theoretical and practical findings:

- 1) initial evidence for a 5-factor HX model [9];
- 2) enriched description of the HX model’s constructs; and
- 3) practical guidelines for quantitatively measuring HX.

II. RELATED WORK

We explore existing UX models and measures, their application to haptics, and other related scales.

A. Haptics and UX

There exist several instruments that evaluate UX, typically collecting potential factors into two dimensions: pragmatic quality and hedonic quality. Pragmatic quality is judged by the practical, goal-oriented aspect of a product, and the efficiency and effectiveness of the users achieving their goals [11]. Hedonic quality is judged by non-tasked quality aspects of the product, such as the aesthetics of the user interface or originality of the design [8]. With the User Experience Questionnaire (UEQ) model [12], the pragmatic dimension is broken down into 3 sub-aspects – Perspicuity, Efficiency and Dependability, and the hedonic quality is broken down into two sub-aspects – Identity and Stimulation. The scales of the AttrakDiff2 [8] questionnaire offers a similar breakdown of UX evaluation, and there is a relatively high correlation between the UEQ scales and AttrakDiff2 scales. Both scales have a third dimension, attractiveness, an overall score for the product combining ergonomic and hedonic quality.

Another UX evaluation framework, the meCUE questionnaire [13] based on the Component model of User Experience, evaluates UX based on instrumental and non-instrumental product qualities. This model aligns with the dimensions of the UEQ model [12] and AttrakDiff2 [8] model, where instrumental qualities resemble pragmatic qualities

*This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC)

¹Suji Sathiyamurthy, Melody Liu, and Oliver Schneider are with the Department of Management Sciences, Faculty of Engineering, University of Waterloo, Waterloo, ON N2L 3G1, CA {snsathiy,wclui,oliver.schneider}@uwaterloo.ca

²Erin Kim is with the Department of Psychology, Faculty of Science, University of Waterloo, Waterloo, ON N2L 3G1, CA erin.kim@uwaterloo.ca

(usefulness, usability), and non-instrumental qualities correspond to hedonic (aesthetics, status, commitment).

However, these UX models are unable to capture factors unique to HX. Haptic experiences are highly dependent on the context of the interaction and feedback from other human senses, such as sound and visuals. Moreover, UX instruments measure the time before, during, and after an interaction, while HX focuses on the experience at the moment of touch.

The HX model [9] defines haptic experience as “*a distinct set of quality criteria combining usability requirements and experiential dimensions that are the most important considerations for people interacting with technology that involves one or more perceived senses of touch, possibly as part of a multisensory experience.*” The model focuses on haptic technology as part of a system, rather than the technology independently. It proposes design parameters, usability requirements, and experiential factors specific for haptics. In this work, we focus on the experiential factors: Harmony, Expressivity, Autotelics, Immersion, and Realism.

B. Other scales that measure related constructs

Other scales related to HX include the Need for Touch (NFT) Scale [14]. NFT measures the user preference of extracting information obtained through haptics experience. The scale focuses on two dimensions similar to the UX model – instrumental (purpose-driven) touch, and the autotelic (hedonic-oriented) factor. This scale is used primarily in market research with non-interactive objects.

Haptic feedback and physical props can increase presence in virtual environments [4]. The presence questionnaires (Presence Questionnaire (PQ) [15] and Immersive Tendencies Questionnaire (ITQ) offer scales overlapping with the Immersion and Realism dimensions of factors of HX.

Haptic feedback is often applied to gaming, and there exists several instruments to measure game user experiences. The Game Experience Questionnaire measures video game-playing effects through the degree of engagement, but was never validated [16]. The player traits model proposes 5 dimensions - aesthetic orientation, narrative orientation, goal orientation, social orientation and challenge orientation [17]. The Player Experience Inventory (PXI) [18] measures how lower-level game design choices directly impact player’s enjoyment. The scale measures constructs such as immersion, mastery, autonomy and functional consequence, has been used to measure how vibrotactile embellishments can increase player experience in mobile games [2].

III. APPROACH

We followed the method of scale development [10], [19] to create, refine, and evaluate a set of items (i.e., questions) to measure HX. Figure 1 outlines this process. While it would be premature to release a draft questionnaire before validation, the insights we have gained so far can inform hapticians about quantitatively measuring UX with haptics, and can invite feedback from the community.

We began with four studies for initial item (question) generation. The first, *1. Scale Inception*, yielded an initial set of



Fig. 1. Scale Development process: 1) Brainstorming provided an initial set of items based on the literature. 2) A group discussion and 3) three group interviews with expert hapticians refined items. 4) Feedback from the target population helped improve understandability. 5) We iteratively piloted for 6) a large online study, which had participants evaluate haptic experiences with our draft scale. 7) Exploratory Factor Analysis (EFA) uncovered the underlying structure of a relatively large set of variables.

25 items. We then iteratively developed these items through 3 studies: 2. *Face Validity* (N=8) for novice evaluation of the items built based on the theoretical model, 3. *Content Validity* (N=6) for expert review of the items’ Relevance and Clarity, and 4. *Cognitive Interviews* (N=9) to evaluate whether the questions are interpretable by the target population using probing questions [20] and think-aloud strategies. We then conducted a 5. *Pilot Study* (N=25) to test our procedure and adequacy of data collection for performing factor Analysis. Finally, we ran our 6. *Survey Administration* (N=261 after cleaning) and conducted exploratory factor analysis (EFA) to study the correlations between the items and the emergent factors. This is the first report of all studies.

We originally hoped to involve a variety of newly commercially available haptic feedback devices, including vibrotactile feedback, variable friction displays, mid-air haptics, and force-feedback. Unfortunately, this research was conducted during the COVID-19 pandemic, and in-person research protocols were suspended. As such, we conducted all steps from 3. *Content Validity* onward remotely, and scoped our primary data collection in the 5. *Pilot Study* and 6. *Survey Administration* steps to only include experiences with smartphones and gaming consoles; i.e., vibrotactile feedback in devices with mass adoption suitable for a large online study. While there is evidence that remote haptic studies with vibrotactile content produce similar feedback to in-lab studies [21], this does reduce the generality of the scale.

IV. RESULTS

We report first on results from our iterative process, then from EFA. Due to space constraints and a desire to not present a draft questionnaire until scale development is finished, we do not report item iteration in detail in this paper.

A. Insights from iterative scale development

The first 3 user studies – face validity, content validity and cognitive interviews – were conducted iteratively. We received insights at each point in the process and ended with a draft questionnaire composed of 22 items divided across each of the 5 experiential factors. For example, at the end of 3: *Content validity*, expert feedback gave strong evidence

that the phrase “Haptic feedback” needed to be included in the items. However, this may make the questionnaire difficult to understand for participants unaware of its definition. We formulated the following definition of haptic feedback and included it with our draft questionnaire:

“Haptic feedback refers to anything that you feel with the sense of touch. It could be vibrations, force, temperature, pressure, or any other physical sensation.”

In step 4 (cognitive interviews), we were able to confirm that including a definition for haptic feedback enhanced the understanding of the items. At the end of this study we narrowed down the applications to use for our survey. Users had two options to choose from for completing a task in which they experience haptic feedback: smartphones and gaming consoles (e.g., Nintendo Switch, PS4, other haptic-enabled consoles). Participants were further presented with a list of games or applications based on the device (e.g., Animal Crossing, Grand Theft Auto) that we carefully selected and created guided tasks for. Participants who selected gaming consoles were also given the option to select a game of their choice which they believed had haptic feedback. We balanced the number of participants using smartphones vs. consoles by restricting the maximum responses for each.

B. Exploratory factor analysis

We deployed a remote online study using Qualtrics [22] and Prolific [23]. The data collected were again from two categories of devices - smartphones and gaming consoles. All the applications had vibrotactile feedback as part of the experience and a total of 302 responses were recorded. 2 of the participants did not complete the study; 261 out of 300 completed responses were identified and verified to have experienced haptic feedback before completing our scale. These participants confirmed that the haptic feedback was noticeable in the exit survey. Items were presented in a random order. The exit survey collected participant satisfaction with haptic feedback, whether they understood the intention of the study, and demographic information: age groups, self-report gender, and familiarity with the haptic device, application, and haptic technology. Questions in the section were used to assess the quality of responses and used in the data cleaning process.

We then conducted factor analysis, a statistical method used to study the dimensionality of a set of variables. It identifies categories of similar statements to model the interrelationships among items. The overall KMO¹ value for our data is 0.93; all observed variables show values greater than 0.90. This suggests that the collected data is adequate for factor analysis. We determined the number of factors that the principal component analysis (PCA) suggests with a scree plot (Figure 2). The initial scree plot using parallel analysis suggested between 3 to 5 factors: each point demonstrates a possible knee. Statistical analysis and interpretation were performed using the psych package in R.

¹A Kaiser-Meyer-Olkin (KMO) [24] test is used to determine the sampling adequacy of data that are to be used for factor analysis.

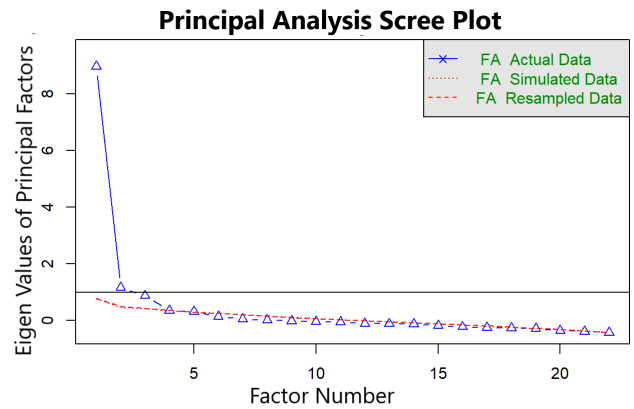


Fig. 2. Scree plot showing the number of possible factors in the data. A drop at the 3rd point and a more subtle drop at 5 indicates that there could be 3 - 5 factors underlying the 22 variables in the data.

The aim of EFA is to find a simple solution where each factor has a small number of large loadings and a large number of 0 (or small) loadings. Starting from 3 factor model (as suggested by PCA), we increased the number of factors and tried different methods and rotations to obtain the simple factor solution. For each possible factor number (3, 4, or 5 factors), we tried Maximum likelihood (ML), Ordinary least squares (OLSS), and Principal Axis Factoring (PAF) extraction methods [25] with both oblique and orthogonal rotations. By trying various combinations of these, we arrived at 36 possible models (we also considered some 6-factor models and varimax-orthogonal rotations) and shortlisted 8 for further evaluation. The goal is to find a model with a strong factor structure, i.e., minimum cross-loading, high factor loading, and sum of squared loadings related to each factor. Based on the above-mentioned metrics, we selected the best model based on highest factor loadings, minimal cross-loadings and other goodness-of-fit metrics. The final model had 5 factors, no cross loadings, all factor loadings $>|0.4|$, TLI and CFI >0.90 and RMSEA and RMS values <0.05 , and had an oblique rotation performed (suggesting correlation between the factors). All 3- and 4-factor models were rejected based on high cross-loadings or failing the goodness-of-fit tests (i.e., RMSEA and RMS values >0.05 & TLI and CFI <0.90).

Figure 3 shows the best performing model. The factor correlation matrix calculated from the data show that all the factors have a correlation of less than 0.7, confirming discriminant validity [26]. 20 of the 22 observed variables had significant loadings onto 5 factors (ML1, ML2, ML3, ML4, ML5), each roughly corresponding to one of the 5 theoretical factors in the HX model. Figure 3 shows the loadings. In this model, “Expressivity” and “Realism” are composed only from items intended to represent those factors. “Autotelic”, “Immersion” and “Harmony” have additional properties loaded onto it along with its other theorized properties. Upon further investigation, we were able to see how these new properties could be associated with the factors that it loaded on and were able to deem the model acceptable.



Fig. 3. Path diagram representing the 5 oblique factors of the final model. The factors were transformed to an oblique solution using promax. Dotted red lines mean a negative loading due to the inverted item. Items are labelled according to the HX dimension we intended them to measure (e.g., “H2” was the second item drafted to represent Harmony); some loaded on a different factor (e.g., “H2” loaded on Autotelic). “H1” and “R4” did not load.

All the signs within “Harmony” were flipped and thus its correlation with other factors were negative. We could either flip all the signs and use it as is or we can re-name the factor to mean the opposite (e.g., Dissonance). We decided to keep the factor as Harmony to be more consistent with the HX model [9], with the requirement that item and factor values need to be negated during measurement or interpretation.

The final 5 factors are correlated with each other: the lowest correlation was 0.4 between Immersion - Expressivity and Expressivity - Autotelics, and the highest correlation was 0.7 between Harmony - Autotelics, Autotelics - Immersion and Harmony - Realism. We can see from Figure 3 that most of our factor loadings are between 0.5 to 0.8 with one factor loading at 0.44. High correlation between factors is not ideal but best practices dictate that the value should not exceed 0.7 (49% shared variance) [26]. This indicates strong item reliability. The negative sign on the factor loadings are that of “negatively-worded” items in the questionnaire and are expected to negatively correlate with the factor. In total, the extracted factors explain 54.7% of the variance.

V. DISCUSSION

Our findings provide support and increased detail for the 5-factor HX model, and guidelines for measuring HX.

A. Support & Elaboration for the HX Model

The results from EFA closely resemble the proposed theoretical HX model [9]: the best-performing model had 5 factors, each of which closely relates to one HX factor. This suggests that the underlying factors of the HX model are distinct (though related) entities that can be measured, and was best modeled with 5 factors. While there might be concerns that this was a “re-discovery” of the 5 factors we used to design the items, best practices in scale development suggest starting with a strong theoretical basis [10].

Autotelic: ML1 turned out to be the strongest factor with maximum influence on its variables. All variables of *Autotelic* (A1, A2, A3, A4) loaded significantly onto ML1 along with another variable from harmony (H2). The high factor loading of all the items is an indicator of how important it is to design a feedback that is likable and satisfying to the user from the sensation alone. The Harmony item H2, questions the satisfaction or desirability of the haptic feedback in the context of the system and has a very high factor loading of 0.805. While Autotelic by definition means that the feedback feels good in and by itself, the factor loadings indicate that the context of the system might significantly influence the likability of the feedback. These results suggest that, for an application to have a good autotelic experience, it necessarily requires context and cannot be considered in isolation.

Expressivity: ML2 consisted of four variables designed for *Expressivity* (*E1, E2, E4, E5*), but *E3* (“The haptic feedback helped me distinguish what was going on”) loaded on Immersion, suggesting that Expressivity has more to do with variance of feedback than causality, and that feedback of someone’s actions is connected to Immersion. We thus suggest that Immersion “allows users to feel their input make an impact on the feedback received” [9], not Expressivity.

Immersion: All positive items of *Immersion* (*I2, I3, I4*) loaded significantly on ML3, confirming our existing understanding of the dimension. We conclude that increased involvement, focus, and engagement is an indicator of an immersive experience. Additionally, as mentioned in Expressivity, *E3* loaded onto this Immersion. This lends support to the idea that when users can both affect and be affected by the system, they become more immersed in it [15].

Realism: Factor ML4 consisted only of variables for *Realism* (*R1, R2, R3*), with the item *R4* (“The haptic feedback matched my expectation”) not loading significantly. This could simply be because the item is not capturing the construct. Since the final inventory can have fewer questions per factor depending on their correlation [19], we can either exclude this item or revise it. Our initial screening had some evidence that participants had some reservations about this question as it requires them to have expectations about the experience as a prerequisite – expectations may not be a good way to detect realism. Another interesting question raised during content validity discussion was about realism vs believability. Some of the experts argued that believability might be a construct of its own and some agreed that it is mostly an aspect of realism. Since *R2* (“The haptic feedback was believable”) has a high factor loading of 0.743, we suggest that the factor “realism” and “believability” are closely linked, since “believability” varies with realism. However, we do not yet have enough information to suggest which might be primary. As “realistic” is a higher loading item compared to “believability” and since it’s consistent with the HX model, we decided to label this factor as “realism”

Harmony: ML5 was originally loaded with 3 negative items (*H3, H5, I1*) and 1 positive item (*H4*). All the negative items were positively correlated with the factor and the positive one was negatively correlated. This factor was in fact the opposite of the construct *Harmony*. We decided to label this factor as harmony, since it simply appears to be dominated by “negative factors”, and can be interpreted with all signs flipped. In other words, Harmony might be the absence of a disruptive feedback. For example, if the haptic feedback does not feel “disconnected” from the system or “out of place,” it may be harmonious.

B. Practical Matters for Measuring HX

1) *A questionnaire is viable for measuring HX:* We repeatedly debated whether a questionnaire was the right choice for evaluating HX. Measuring experience is difficult. Focusing on a particular type of feedback while keeping in mind the overall experience is even more challenging. Through brainstorming and reviewing work around measure-

ment instruments, we were convinced that a well-designed questionnaire can collect systematic data of user emotions and expectations about a system or design. We found participants typically interacted with haptic experience for 5 minutes, with the entire task and questionnaire taking 10 minutes. This suggests that participants can complete a 20-question instrument in less than 5 minutes, suitable for evaluating systems without being too onerous on participants.

However, approximately 12% of our survey participants reported that they did not feel any haptic feedback in the guided task and their responses consisted only of neutral options. Measurement of HX may only be applicable for applications with a prominent haptic feedback component, or when participants direct their attention to the haptics.

2) *“Haptic feedback” needs to be defined for general participants:* We initially referred to the unit measured as “feedback”, in order to keep it generic and simple. However, “feedback” was ambiguous as participants were confused about what feedback we were referring to. We then decided to change it to “haptic feedback” based on the input from experts. This change could potentially make the questionnaire difficult to comprehend for novice participants: 57.7% of our survey participants had not heard of the word or unsure of what it means. Thus, we included a simple definition of haptic feedback (as described in Results). We found that defining haptic feedback was necessary to help users understand the items from 82.9% of survey respondents.

3) *Not all factors apply in all contexts:* In order to build our studies around measuring haptics, we had to provide participants with applications that are experienced as a system, but also have recognizable haptic feedback. For example, watching a haptic ad - which is very similar to watching any video with sounds and visuals. We conducted an initial screening with smartwatches and fitness tracking devices and discovered that not all our constructs may apply to simpler applications, such as a timer or a “step goal reached” alert. Some applications don’t have a real-world equivalent, so “believability”, “realistic” and “convincing” seem irrelevant to the experience, or confusing; one participant interpreted “believable” as believing that the vibration came from their FitBit. Thus, any resulting questionnaire should ideally be designed with sub scales that can be selected depending on the needs of the researcher and the context of the study.

VI. LIMITATIONS

While there are promising results from the study, there are several limitations. Game-like applications were used in the study to measure haptic feedback; other types of haptic applications may not perform the same. It is possible that users associate games with pleasant experiences and therefore lean towards giving higher scores. There could also be social desirability bias, a systematic error in self-reporting measures resulting from participant desire to avoid embarrassment and project a favorable image to others [27]. Future validation will help us identify, test and prevent bias.

Longitudinal and repeat studies will be necessary for validation of these results and any instrument. The quan-

titative studies we conducted are cross-sectional studies, i.e., a snapshot of the population about which the data was collected, and do not consider the effect of time.

Due to the remote nature of the main study, we were not able to interact with our users as freely as in an in-person study, and relied on the messaging functionality offered by the crowd-sourcing tool (which helped resolve some of the issues and questions participants had). We were only able to use commercially available devices with vibrotactile feedback, since finding remote participants with advanced setups (and controlling them) was very tedious. We expect our findings will translate to in-person studies, as prior work suggests remote vibrotactile studies on commodity hardware yield similar findings to in-person studies [21].

Finally, we acknowledge the risks of publishing a non-validated scale. The developed items are not ready for deployment as a questionnaire, and **we do not recommend using these questions until a validated scale is produced.**

VII. CONCLUSION

In this paper, we used scale development to elaborate on the HX model. We found that a 5-factor model was the best performing model, lending support to a 5-factor theoretical construct. Our analysis provides more description of each experiential dimension and recommendations on how to use questionnaires to measure HX. These findings will help hapticians inform and evaluate their designs, and push towards better haptic experiences. Our next steps are to work towards validating our findings, expand to other modalities, try separate subscales, and consider the impact of time.

VIII. ACKNOWLEDGEMENTS

We thank the reviewers for their comments. This work was supported by Canada's NSERC Discovery Grant program.

REFERENCES

- [1] E. Maggioni, E. Agostinelli, and M. Obrist, "Measuring the added value of haptic feedback," *2017 9th International Conference on Quality of Multimedia Experience, QoMEX 2017*, pp. 1–6, 2017.
- [2] T. Singhal and O. Schneider, "Juicy haptic design: Vibrotactile embellishments can improve player experience in games," in *CHI '21 (in press)*, 2021, p. 10.
- [3] H. Pauna, P.-m. Léger, S. Sénécal, M. Fredette, F. Courtemanche, S.-L. Chen, Labonté-Lemoyne, and J.-F. Ménard, "The psychophysiological effect of a vibro-kinetic movie experience: the case of the D-BOX movie seat," in *Information Systems and Neuroscience*. Springer, 2018, pp. 1–7.
- [4] M. Azmandian, M. Hancock, H. Benko, E. Ofek, and A. D. Wilson, "Haptic Retargeting," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5 2016, pp. 1968–1979. [Online]. Available: <https://dl.acm.org/doi/10.1145/2858036.2858226>
- [5] C. C. Berger, M. Gonzalez-Franco, E. Ofek, and K. Hinckley, "The uncanny valley of haptics," *Science Robotics*, vol. 3, no. 17, pp. Art—No, 2018.
- [6] O. Schneider, K. MacLean, C. Swindells, and K. Booth, "Haptic experience design: What hapticians do and where they need help," *International Journal of Human Computer Studies*, vol. 107, no. August 2016, pp. 5–21, 2017. [Online]. Available: <http://dx.doi.org/10.1016/j.ijhcs.2017.04.004>
- [7] J. Lessiter, J. Freeman, E. Keogh, J. Davidoff, V. Zamanzadeh, A. Ghahramanian, M. Rassouli, A. Abbaszadeh, H. Alavi-Majd, A.-R. Nikanfar, User Interface Design GmbH, S. Kujala, V. Roto, K. Väänänen-Vainio-Mattila, E. Karapanos, and A. Sinnelä, "AttrakDiff," *Presence: Teleoperators and Virtual Environments*, vol. 10, no. 3, pp. 282–297, 6 2001.
- [8] M. Hassenzahl, M. Burmester, and F. Koller, "AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität," 2003, pp. 187–196. [Online]. Available: http://link.springer.com/10.1007/978-3-322-80058-9_19
- [9] E. Kim and O. Schneider, "Defining Haptic Experience: Foundations for Understanding, Communicating, and Evaluating HX," in *CHI 20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 5 2020. [Online]. Available: <https://doi.org/10.1145/3313831.3376280>
- [10] G. O. Boateng, T. B. Neilands, E. A. Frongillo, H. R. Melgar-Quinonez, and S. L. Young, "Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer," *Frontiers in Public Health*, vol. 6, no. June, pp. 1–18, 2018.
- [11] M. Hassenzahl, "The Hedonic/Pragmatic Model of User Experience," in *Towards a UX Manifesto. COST294-MAUSE affiliated workshop*, 2007.
- [12] B. Laugwitz, T. Held, and M. Schrepp, "Construction and Evaluation of a User Experience Questionnaire," 2008, pp. 63–76. [Online]. Available: http://link.springer.com/10.1007/978-3-540-89350-9_6
- [13] M. Minge, M. Thüring, and I. Wagner, "Developing and Validating an English Version of the mcCUE Questionnaire for Measuring User Experience," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 60, no. 1, pp. 2063–2067, 9 2016. [Online]. Available: <https://doi.org/10.1177/1541931216601468>
- [14] J. Peck and T. L. Childers, "Individual Differences in Haptic Information Processing: The "Need for Touch" Scale," *Journal of Consumer Research*, vol. 30, no. 3, pp. 430–442, 12 2003. [Online]. Available: <https://doi.org/10.1086/378619>
- [15] B. G. Witmer and M. J. Singer, "Measuring Presence in Virtual Environments: A Presence Questionnaire," *Presence: Teleoperators and Virtual Environments*, vol. 7, no. 3, pp. 225–240, 6 1998. [Online]. Available: <https://doi.org/10.1162/105474698565686>
- [16] E. L.-C. Law, F. Brühlmann, and E. D. Mekler, "Systematic Review and Validation of the Game Experience Questionnaire (GEQ) - Implications for Citation and Reporting Practice," in *The Annual Symposium on Computer-Human Interaction in Play Extended Abstracts - CHI PLAY '18*, 2018, pp. 257–270. [Online]. Available: <http://dl.acm.org/citation.cfm?doid=3242671.3242683>
- [17] G. F. Tondello, K. Arrambide, G. Ribeiro, A. J.-I. Cen, and L. E. Nacke, "'I Don't Fit into a Single Type': A Trait Model and Scale of Game Playing Preferences," 2019, pp. 375–395. [Online]. Available: http://link.springer.com/10.1007/978-3-030-29384-0_23
- [18] V. Vanden, K. Spiel, L. Nacke, D. Johnson, and K. Gerling, "A scale to measure player experiences at the level of functional and psychosocial consequences," *Journal of Human Computer Studies*, vol. 135, no. October 2019, p. 102370, 2020. [Online]. Available: <https://doi.org/10.1016/j.ijhcs.2019.102370>
- [19] R. F. DeVellis, "Scale Development Theory and Applications (Fourth Edition Robert)," *SAGE Publication*, vol. 4, p. 256, 2016.
- [20] H. Haeger, A. D. Lambert, J. Kinzie, and J. Gieser, "Cognitive Interviews to Improve Survey Instruments," *Annual Forum of the Association for Institutional Research*, no. June, p. 15, 2012.
- [21] O. S. Schneider, H. Seifi, S. Kashani, M. Chun, and K. E. MacLean, "HapTurk: Crowdsourcing Affective Ratings for Vibrotactile Icons," in *CHI '16*. New York, New York, USA: ACM Press, 5 2016, pp. 3248–3260. [Online]. Available: <https://doi.org/10.1145/2858036.2858279>
- [22] "Qualtrics XM // The Leading Experience Management Software." [Online]. Available: <https://www.qualtrics.com/>
- [23] "Prolific — Online participant recruitment for surveys and market research." [Online]. Available: <https://www.prolific.co/>
- [24] "Kaiser-Meyer-Olkin (KMO) Test for Sampling Adequacy - Statistics How To." [Online]. Available: <https://www.statisticshowto.com/kaiser-meyer-olkin/>
- [25] J. Osborne, J. W. Osborne, A. B. Costello, and J. T. Kellow, *Best Practices in Exploratory Factor Analysis*, 2011, no. January 2005.
- [26] "StatWiki." [Online]. Available: http://statwiki.kolobkreations.com/index.php?title=Main_Page#Discriminant_validity
- [27] R. J. Fisher, "Social Desirability Bias and the Validity of Indirect Questioning," *Journal of Consumer Research*, vol. 20, no. 2, p. 303, 9 1993.